The ethics of the ethics of autonomous vehicles: Levinas and naked streets

Abstract

My starting point in this paper is that investigating the ethics of autonomous vehicles through the lens of the trolley problem is not only limited but also unethical. I construct my case by aligning myself with Niklas Toivakainen, who argues against David Gunkel's reading of Levinasian ethics as an answer to the 'Machine Question'. I adumbrate Toivakainen' s critique that the attempt to give a Levinasian face to the machine is an example of a compensatory logic – a way to avoid a deeper exploration of the moral dynamics from which our technologies emerge. I offer an extended reading of Levinas's formulation of the face of the other to argue that while the machine cannot signify as a face, the machine can announce or anticipate the signification of the face which the self is infinitely responsible to. I reach this conclusion through focusing on the uncertainty and ambiguity of the signification of the face. I then argue that current approaches to the ethics of autonomous vehicles, based on variations of the trolley problem, fail to locate ethical responsibility to the other, precisely in its attempts to evade uncertainty. I demonstrate how an embrace of uncertainty can make the ethics of autonomous vehicles more ethical through a consideration of the concept of shared space, and naked streets, in the work of the late Dutch traffic engineer Hans Monderman.

Autonomous vehicles are one of the more exciting and disruptive technologies coming out of the fourth industrial revolution (4IR). In response, a thriving cottage-industry has emerged to tackle the vexatious issue of the ethics of self-driving vehicles. Who, it is asked, should be held responsible for injuries and fatalities that might arise from any collision in, or with, an autonomous vehicle? Should the owner of the vehicle, or the manufacturer be held responsible? (Hevelke and Nida-Rümelin 2014) Or should the engineers who design the algorithms that steer the vehicle be held responsible? (Nyholm and Smids 2016) In attempting to answer these questions, the ethics of autonomous vehicles seems to have settled on the so-called

'trolley-problem' as the lens through which to explore solutions. The trolley problem[1] (whose genesis goes back to a paper by Phillipa Foot (1967)) imagines a scenario in which a moral agent has the power to prevent a run-away trolley from colliding into, and killing, five persons, by pulling a lever which diverts the trolley onto another track only to collide with, and kill, one other person.

The ethics of autonomous vehicles employs these utilitarian calculations to determine, inter alia, how an autonomous vehicle should react to one or more persons, or children, or animals suddenly, and unexpectedly, appearing in its path. Should it swerve violently to avoid a pregnant woman, but not an old man, and risk death or injury to the driver and/or passengers?[2] Some have suggested that autonomous vehicles might incorporate 'personal ethics settings' (not unlike interior climate control settings) which would allow the owners of the autonomous vehicles themselves to decide which life they place a greater value on – pregnant women or elderly men (Gogoll & Muller 2017). Other theorists, like Himmelreich (2018), warn against the siren of fanciful, though not improbable, scenarios that form the basis of many trolley problems, urging us instead, to repurpose the title of his paper, to mind the trolley and consider the ethics of autonomous vehicles in mundane situations. Himmelreich (670) argues that "the usefulness of trolley cases in investigations of the ethics of autonomous vehicles is limited", joining authors such as Goodall (2016), and Nyholm and Smids (2016).

[1] Himmelreich (2018) distinguishes between trolley *cases* and trolley *problems* – the former consisting of a scenario in which there exists only the option between colliding with a person appearing in front of a car or swerving to avoid a collision with that person; while the latter consists of two or more cases taken together. Himmelreich notes that many contributions in the literature do not make this distinction. I will not make this distinction as it is not pertinent to my argument.

[2] See Awad et al. (2018) in which participants in a world-wide survey on autonomous vehicles indicated that they would more likely favour evasive action that saved the life of a pregnant woman than the life of an old man.

In this paper, I will argue that the usefulness of trolley cases in interrogating the ethics of autonomous vehicles is not only limited, but also unethical. In exploring what the ethics of the ethics of autonomous vehicles encompasses, I will then be circumscribing a meta-ethical position[3]. The ethics of autonomous vehicles, I claim, cannot be considered apart from what it means to be, in the terms used by Emmanuel Levinas (1969), "infinitely responsible to and for the other".

I will characterise all formulations of the trolley problem as an example of what Toivakainen (2016) calls 'compensatory logic' – a manner of thinking which obscures whether progress (human advancement) stems from the ethical meaning we make of (technological) progress, or whether "ethics is forced to accommodate itself to the demands of progress" (279). Toivakainen illustrates his notion of compensatory logic by examining companion robots in elderly care. He then asks, 'what is our relationship to machines really like?', which he later clarifies in the following way:

> How it is that we have come to a situation in which we need to start developing care-bots in the first place – and what kinds of moral dynamics is involved here? (*ibid.*)

My project directs Toivakainen' s question toward autonomous vehicles and asks what moral dynamics are involved if we consider that the primary reason given for developing autonomous vehicles is improved road safety. Toivakainen, following

[3] A meta-ethical claim in the standard sense of the term typically takes an ethical claim as its subject matter but is not itself an ethical claim. I wish to thank an anonymous reviewer for drawing my attention to this point. In this paper I diverge from this practice which is necessitated by the distinction I draw between ethics and *Levinasian* ethics. At this preliminary stage one way to note the distinction is that 'ethics' attaches to *theory* and involves principles, norms and imperatives. Levinasian ethics eschews any overarching theory and does not posit any ethical principles. This will be made clearer in the paper's first section. Kantianism, Utilitarianism and Virtue ethics constitute the triumvirate of standard contemporary ethical theories. In the context of the trolley problem and autonomous vehicles, the term 'ethics' should be disambiguated even further because these are invariably discussed from within a utilitarian framework.

Gunkel (2012), enlists the ethics of Emmanuel Levinas, who argues for ‘ethics as first philosophy’, to claim that we find ourselves always already enmeshed *within* ethics. Every situation is normatively constituted. This means that the ethics of autonomous vehicles, or companion robots, cannot be considered after the technological details are worked out. Gunkel (2012) applies Levinasian ethics in order to posit a ‘machine-face’, which would allow moral consideration to extend to machines, including autonomous cars and care-bots. Toivakainen disagrees with Gunkel’s move, arguing that a machine cannot have, or be, a face in the Levinasian sense. I side with Toivakainen.

I will also enlist the ethics of Emmanuel Levinas (1969) but will emphasise different aspects of his thought to make my case. In particular, I will focus on how the face of the other signifies its ethical singularity as “meaning all by itself” (Levinas 1985; 86). Toivakainen does reference this phrase, but in passing, and assigns it a very narrow import. I will expand on other Levinasian formulations of the face, inter alia, that the face is “signification without context” *(ibid*.).

To illustrate how these ideas might be brought to bear on practice, particularly within the context of road safety, I will turn to the work of the late Dutch traffic engineer Hans Monderman. Monderman’ s conceptualisation of ‘naked streets’, implemented in various forms in over 400 towns across Europe, sees road markings and other traffic signals reduced to a bare minimum (Jenkins 2016). Removing traffic signs creates ambiguity, thus inducing uncertainty in the driver. This in turn heightens the driver’s responsiveness, which, inter alia, reduces speeding, making the road safer for all users. I make the case that traffic signals and roads, as a form of technology, can also signify the face of the other. More specifically, traffics signals and roads, in announcing or anticipating the face of the other, signify the singularity and

vulnerability of the other to whom I am infinitely responsible[4]. The conclusion I reach is that in embracing the ambiguity and uncertainty that sharing public space entails, particularly on the road, the ethics of autonomous cars can be made more ethical.

The paper will proceed as follows: In the section to follow, I trace the critique that information ethics (IE) follows on from a centrist driven ethics that seeks to expand moral consideration from a privileged (human) 'we' outwards to include machines. The trolley problem, framed within Utilitarianism, is paradigmatic of the dilemmas of responsibility centrist driven ethics gives rise to within the ethics of autonomous vehicles. Gunkel employs Levinasian ethics to address centrist ethics' shortcomings within the 'machine question'. Toivakainen charges Gunkel with a misreading of Levinas. In section two I adumbrate Toivakainen 's notion of a compensatory logic which prevents us from delving deeper into questioning our relationships with machines, moving from his example of companion robots to this papers' central concern with autonomous vehicles. Compensatory logic forestalls an interrogation of the moral dynamics which frame the emergence of any particular new technology and their impact on society. In the present case, compensatory logic forestalls an ethical consideration of the impact of autonomous vehicles on road-safety when ethical consideration is understood in Levinasian terms. In section three I expand on how the Levinasian face signifies the alterity, or 'otherness' of the other as 'meaning all by itself'. This is in service of supporting the claim that the machine has no alterity of itself that is capable of signification. Instead, what the machine can signify is the alterity of the other person and the responsibility that is owed to her. In the final section of the paper, I turn to the example of Monderman' s 'naked streets' in order to illustrate how the ethical responsibility owed to the (Levinasian) face might be

[4] I wish to thank an anonymous reviewer for helping me clarify this important claim.

signified within, or through certain technologies, as against those technologies being able to signify meaning ‘all by themselves’ the way a (Levinasian) face does. In their privileging of the Other, these considerations offer a more ethical approach than the trolley problem to developing the ethics of autonomous cars.

Levinasian ethics: An ethics of ethics

Centrist ethics and the machine question

In ‘Machines and the face of ethics’, Niklas Toivakainen (2016) critically assesses David Gunkels’ response, in *The Machine Question* (2012), to the exclusivist or centrist logic that undergirds most IE. Centrist ethics operates by positing a privileged ‘we’ whose shared features, such as rationality, constitute grounds for moral consideration. This central ‘we’ is then expanded by demonstrating how a particular attribute “is also found in the ‘other’ (or a segment of them) or then to identify some new characteristic upon which the new ‘we’ gets founded” (Toivakainen 2016; 270). So, for example, utilitarians bring non-human animals into the circle of moral consideration by arguing that they share a capacity for suffering with humans. Similarly, IE seeks to broaden the moral ‘we’ by arguing that “anything that is a coherent body of information ought to be worthy of our moral consideration” (*ibid.*). As such, this will allow space for robots and machines within the ethical domain, either as moral agents or moral patients, or both.

The charge against centrist ethics is that by pursuing sameness it necessarily elides what is different and singular (271). To qualify for entrance into the newly expanded ‘we’, the excluded other must become like the ‘we’. If we cannot recognise some feature in the other, the basis on which we can say that the other is *like us*, then we have insufficient reason to broaden the ethical domain to include that other. To

counter this reductionism and as a solution to 'the machine question', Gunkel (2012) enlists the ethics of Emmanuel Levinas. Levinasian ethics is not merely a privileging of otherness, or what Levinas terms 'alterity', over sameness; otherness is what constitutes subjectivity (Levinas 1969). More particularly, as Toivakainen (2016; 271) clarifies, it is the

> *relationship* between the self and the other [which] is in fact the constitutive foundation of the very structure of the subject and of ethics.

The relationship between the self and the other is one marked by a radical asymmetry: the autonomy of the self is held hostage to the ethical demand the vulnerability of the other before me creates – an unconditional, non-reciprocal infinite responsibility for her (Levinas 1969; 43). This is Levinas's definition of ethics – the "calling into question of my spontaneity by the presence of the Other" (*ibid.*). Such a definition resists an easy answer to the classic ethical question, 'What ought I to do?', whether it be acting to maximise overall happiness – utilitarianism – or acting so as not to treat the other as a means – Kantianism[5].

Levinas (1985; 99) claims that I owe the other everything –

> I am responsible for a total responsibility, which answers for all the others and for all in the others, even for their responsibility. The I always has one responsibility *more* than all the others.

I can expect no reciprocity from the other, however – "in pure charity, I know what I owe the other. What the other owes me, that's his business (2001; 54). How does Levinas defend this bewildering claim?

[5] See FN.3 above on the distinction between ethics and *Levinasian* ethics, in which the former is generally explicated within an ethical *theory* such as Kantianism or Utilitarianism.

Levinas (1969; 39) argues that “[T]he absolutely other is the Other”. In other words, the other person presents me with a radical otherness, a radical alterity. Any attempt to understand this radical alterity of the other will require me to enlist categories such as gender and race, or reasoning and suffering ability, the very features, the basis on which, centrists seek to broaden a privileged ‘we’. In the process of attempting to represent the ethical demand as X or Y, as a demand for restorative racial justice, for example, I am reducing the others’ radical otherness to sameness. Levinas describes this as a ‘totalisation’ of the other. Bauman (1993; 90) describes how the alterity of the other eludes us, and its mired in ambiguity and uncertainty:

> I embark on the search for the content of the command. But I cannot find that content in any way except through ‘representing’ […] What I ‘find’ is the Other’s command as I have articulated it; my representation of the Other’s voice.

I cannot determine what my specific responsibilities to the other before me are because I cannot determine the other’s singularity. I cannot comprehend the “strangeness” (Levinas 1969; 43) of the other before me, only that I am completely and irrecusably responsible, somehow, to and for her.

The upshot of this is that the self, in Gunkel’s (2012; 177) words,

> does not constitute some pre-existing self-assured condition that is situated before and is the cause of the subsequent relationship with an other. It does not (yet) take the form of an active agent who is able to decide to extend him- or herself to others in a deliberate act of compassion. Rather it becomes what it is as a by- product of an uncontrolled and incomprehensible exposure to the

> face of the Other that takes place prior to any formulation of the self in terms of agency.

For Levinas (1998; 114), the self is not a sovereign totality who constitutes themself in an act of Cartesian self-creation – 'I think, therefore I am' – rather, the self is an election by the other – 'I am chosen, therefore I am'. It is the alterity of the other which the self cannot incorporate into its knowledge of the world which arrests the self's agency, just like the idea of infinity whose ideatum cannot be incorporated into its idea (Levinas 1969; 49). The 'face' is the central figure in this encounter with the other, who transforms the subject from a *for-itself* into a subject *for-the-other*, from one interested in its own freedom to one disinterested.

The term 'face' has a specific meaning within Levinasian ethics and is arguably his best known and celebrated concept[6]. The face of the other is *the mode* in which the other presents her alterity (Levinas 1969; 50) rather than the composite of inter alia, eyes, nose, and mouth. The face is thus better understood as a verb – the face presents the uniqueness of the other through *facing* me in an originary encounter. Levinas (1985; 86) says that the face is "meaning all by itself". Toivakainen' s (2016; 272) reading of this important Levinasian citation is that

> ethics is not something the 'self' can determine […] It is not we that give the other's face its meaning or construct it, but rather, it is through our encounter with the (meaning of the) face of the other that we become (gain the meaning of) a 'self', a self with thoughts, emotions, consciousness etc., a self that then can impose and project all kinds of 'meanings' on the other.

[6] The face takes centre stage in his first magnum opus, *Totality and Infinity* (1969); whereas in his second, *Otherwise than Being* (1998), the concept of the Saying and the Said is foregrounded.

The idea of an originary face to face encounter and the rejection of a 'pre-existing self-assured' subjectivity is captured in Levinas's (1989) claim that ethics is first philosophy. Ethics comes before ontology – who, or what is the other or subject? – and before epistemology – how can we know the other and the other's demands on us, or, how can the subject know itself? Ethics as first philosophy seeks to *justify* our human existence before explaining our existence. Gunkel (2012; 175) expresses ethics as first philosophy as ''a kind of morality before morality''. Toivakainen (2016; 271) elaborates on this and explains that "any ethical theory or norm already presupposes that which makes ethics *ethics.*" Toivakainen (2016; 271) uses Floridi's(2002; 300) normative claim that 'fighting information entropy is the general moral law to be followed' "fighting information entropy is the general moral law to be followed

to illustrate the Levinasian point that this norm by

> *itself* does not clarify what makes it ethical, but always points to something that is always already there, something in relation to which the norm is established, something that makes *it ethical.*

For Levinas, what 'makes something' ethical, whether it be a normative claim or principle, finds its ultimate source in the singularity of the other whom I encounter face to face.

The subtitle of this paper takes its inspiration from Derrida's (1978; 111) formulation of Levinasian ethics as "an ethics of ethics". An ethics of ethics offers itself as a means to evaluate whether any particular way of discovering or determining the ethical is itself an ethical way to proceed. This paper's aim is to demonstrate that the ethics of autonomous vehicles, hitherto variations on the trolley problem, in turn,

explorations framed within Utilitarianism, has not itself been ethical in the Levinasian sense.

Toivakainen (2016; 274) traces how Gunkel attempts to posit a 'machine face' by building on the not uncommon charge that the Levinasian face privileges the human over the non-human (cf. Llewelyn 1991; Diehm 2000; Derrida 2008). While Levinas (1988) was deeply ambivalent on the possibility of an 'animal face', some commentators have read Levinas to grant not only non-human animals a face (Atterton 2011) but also the environment (Casey 2001). In granting animals and the environment a face we can no longer claim that they do not make any moral demands on us. Possessing a (Levinasian) face means that we are responsible to and for animals and the environment because they are not *like us*. Gunkel hopes to replicate this strategy in order to grant the machine ethical status[7]. Gunkel (2012) argues that

> before something is decided to be either a moral agent or a moral patient, we make a decision whether to make this decision or not (175) [… and…] We, individually and in collaboration with each other (and not just those others who we assume are substantially like ourselves), decide who is and who is not part of the moral community (214).

Toivakainen' s critique of Gunkel's reading of Levinas

It is at this point that Toivakainen (2016; 274) parts way with Gunkel, and correctly so. Despite Gunkel's "respectful relation" to Levinasian ethics (as first philosophy),

[7] In 'Is the machine question the same question as the animal question?' Hogan (2017) also critically analyses Gunkel's *The Machine Question* (2012). She argues that the machine and animal questions are not alike, basing her argument on the distinction between moral agency and moral patientcy, a distinction, she argues, Gunkel fundamentally misconstrues.

what Gunkel gets wrong about Levinasian ethics is his claim that it is up to us, whether individually or collectively, to *decide* who (or what) is worthy of moral consideration (*ibid.*). It is not up to us to decide who has a face or not, that is, it is not up to us to decide where our responsibility to and for the other ends. To recall, ethics is, on Levinas's account, an arresting of our autonomy by the other – we are hostage to the demands of the other. We are not free to decide whether to be in an ethical relation with the other or not, that is, we are not free to decide that X counts as a face and Y does not. This does not mean that I cannot turn away from the face and ignore its demands, but that is a matter of proposing norms in which we decide who or what to exclude from moral consideration. I will get to how Levinas regards norms, and normativity, as opposed to ethics, shortly.

Another way to understand this criticism is to consider Toivakainen' s response to Gunkel's (2012; 201) conclusion that "cybernetics" (Floridian IE, *inter alia*) offers one of the more promising ways to offer "a radical, posthuman *theory* that deposes anthropocentrism and opens up thoughtful consideration to previously *excluded* others". Toivakainen (2016; 274) argues that Gunkel's (flawed) "understanding of the 'decision'' involved in [Levinasian] ethics takes him back to an idea that it is up to our decision to *include* the machine". As such, Gunkel succumbs to the very centrist paradigm he so virulently opposes. Toivakainen attributes this to Gunkel's "conception that in order to reach a higher level of inclusion, we are in need of a better *theory*" (*ibid.*).

Levinas does not offer us a *theory* of ethics because theories require categories, concepts, and systemisation, which in turn reduce the alterity of the other into a totalising sameness. This leaves any attempt to 'apply' Levinas to any normative

question, machine or otherwise, on tenuous terrain. Elsewhere, Levinas (1985; 90) will expand on what this entails:

> My task does not consist in constructing ethics; I only try to find its meaning. One can without doubt construct an ethics in function of what I have just said, but this is not my own theme.

By ‘constructing ethics’, Levinas means precisely a theory of ethics, or a system of norms to guide ethical behaviour; while the meaning of ethics is grounded in the relationship with the other in the originary face-to-face encounter.

Commentators in the literature have gone about ‘constructing ethics’ by capitalising on Levinas’s notion of the third party (1969; 212-5) – the other of the other – to the original dyad of self and other. Exploring this important theme is beyond the scope of this paper. However, I will include one citation in which Levinas (1998; 157) explains that the third,

> is of itself the limit of [infinite] responsibility [to and for the other] and the birth of the question: What do I have to do with justice? […] Justice is necessary, that is comparison, coexistence, contemporaneousness, assembling […] ‘Justice’ is for me something which is a calculation, which is knowledge, and which supposes politics; it is inseparable from the political. It is something which I distinguish from ethics, which is primary.

In short, all the formulations of ‘constructing ethics’, or a theory of ethics, or a system of normative injunctions; or dispensing justice and circumscribing politics, amount to the same thing, which Critchley (2004; 178) neatly sums up as “the art of a response to the singular demand of the other”. I will explore one interpretation of this ‘art’ in the

third section of the paper where Levinas talks about the gesture of ‘after you, sir’, as an example of demonstrating infinite responsibility to and for the other.

To recapitulate this section’s salient points: IE, which includes the ethics of autonomous vehicles, has typically approached its problems by employing ‘centrist ethics’ – a positing of a privileged ‘we’ whose shared features with other agents constitute grounds for moral consideration. However, centrist ethics, in pursuing sameness, elides that which is other. A corrective to this is found in Levinasian ethics where otherness, or alterity is constitutive of the subject. The Levinasian subject arises in a face-to-face encounter with the other in which the alterity of the other arrests the agency of the self. Two readings of Levinas within the context of the ‘machine question’ were then outlined. I followed Toivakainen ‘s position that Gunkel misconstrues Levinas insofar as Gunkel claims that it is possible to decide, using ethical *theory,* who (or what) is worthy of moral consideration.

The compensatory logic of the machine-face

Returning to Toivakainen’ s critique of Gunkel, Toivakainen (2016; 275) argues that moral consideration was extended to slaves and animals, inter alia, not because these were given a face –

> the problem was not that the slave did not have a face; the problem was that the slave-society suppressed and repressed it. [...Rather…] their faces have now been acknowledged and permitted to be (socially accepted/acknowledged as) faces.

Toivakainen is now able to clarify the question that Gunkel fails to ask and which results in Gunkel falling back onto a centrist logic:

> Why it is that we should at all be concerned with our moral relationship — or lack of it — to machines? […] Why are we concerned with how to relate to machines? Why does Gunkel feel it to be important to ask about the face of machines? Have we mistreated the machine, the way slaveowners mistreated their slaves, in ways that indicate that it in fact has a face that has been repressed? (*ibid*.).

Rather than asking whether the machine has a face, Toivakainen argues that we should instead ask, "what is our relationship to machines really like?" (*ibid*.). Furthermore, "if the machine is to have a face, then it already has one, or if it is a question of a future machine then similarly it must bear its own ethical meaning and cannot be given one" (*ibid*.). In other words, if machines are entitled to moral consideration it is not because they do not have a face but because we have, wrongly, supressed that face. That the machine must bear its own ethical meaning is in reference to Levinas's claim that the face is 'meaning all by itself' (*op cit*.).

Toivakainen cites this phrase – 'meaning all by itself' – twice (272; 274) in the course of his paper, which he says he takes to mean that "who or what has a face is not up to us to decide". I have examined this understanding above. However, Toivakainen does not consider other formulations Levinas offers of the face's 'meaning all by itself'[8]. In the very same cited passage, Levinas (1985; 86) says that the "face is signification, and signification without context". What this amounts to, argues Levinas (2001; 48), is that the face "is an appeal or an imperative given to your responsibility: to encounter a face is *straightaway* to hear a demand and an order". In seeking to

[8] Toivakainen (2016; 275-278) supplements his Levinasian analysis by critiquing the work of Introna (2014) in a section called 'In search of the machine-face'. Introna elaborates on Heidegger's notion of *Gelassenheit,* 'letting be', in order to include artefacts such as machines into the ethical domain. I will not comment on this as my concern is only with Toivakainen' s notion of 'compensatory logic'.

answer Toivakainen' s question in the context of the ethics of the ethics of autonomous cars, I will focus and expand upon how and what the face signifies, as well as the ambiguity and uncertainty of this signification. I undertake this in the section to follow. I now continue Toivakainen' s critique by explicating his notion of 'compensatory logic'.

Toivakainen (2016; 279) sharpens his question – what is our relationship to machines really like? – further on as

> how it is that we have come to a situation in which we need to start developing [technology 'X'] in the first place – and what kinds of moral dynamics is involved here – when all the resources are focused on 'progress and 'development'?

For technology 'X' Toivakainen discusses the case of companion robots in elder-care, in particular, the interactive seal-doll Paro[9]. I will be discussing autonomous vehicles as another instance of technology 'X'. In the case of companion robots in elder care, the driver of technological advancement is the rationalisation of medical care and the imperative to cut costs, human carers being more expensive (*ibid*.). Animal companions are one tried and tested way to ease the loneliness of residents in elder care homes. However, animal themselves require special care and their housing and feeding is costly. An animal robot as a companion solves these issues. What is absent here is the examination of the *moral dynamics* that drive these calculations. These might include considerations of the dissolution and/or reconfiguration of the traditional extended family; or the impact of an ageing population (especially in developed countries) on public resources. How should we

[9] See http://www.parorobots.com/

allocate resources between the competing demands of primary and elderly healthcare? Or with education? Who should bear the greater responsibility for the elderly? Society or the family? These are all normative considerations that should require our attention *before* embarking on the quest to make a robotic seal more empathetic. Toivakainen (281) remarks that

> robots, to the extent they are meant to take over the task of care work, are an extension of the form of alienation that is already included in institutionalised care work.

Toivakainen (280) calls the quest for a machine-face, or a more empathetic robot, a ‘compensatory logic’, a way to “compensate for absent humans and living beings”. Compensatory logic allows us to avoid asking the tough questions of whether companion robots and other technologies are

> ways of taking responsibility or avoiding it? Do they build on a conception of freedom as something tied to responsibility, or a conception of freedom solely built on rights and independence from others? (*ibid.*).

Toivakainen argues that compensatory logic leads to technological advances that while seeming to contribute “something new and important’, can instead, “obscure a problem by ‘solving’ it without addressing it” (Turkle 2011; 283 in Toivakainen 2016; 280). How might autonomous vehicles obscure a problem by solving it, with personal ethics settings for example, without addressing that problem? To answer that question, I will identify the primary problem that autonomous vehicles purport to solve is traffic safety (Gogoll & Muller 2017).

What the ethics of autonomous vehicles also fails to adequately address is our moral attitudes with respect to the law in general and road-safety law in particular.

Otherwise law-abiding citizens will regularly speed or talk on their cell-phones whilst driving. Why do such infractions of the law attract such blasé attitudes when they are significant contributing factors to road accidents and fatalities? While social mores towards drunk-driving have hardened in the last two decades or so, it used to raise few eye-brows when an inebriated friend insisted, slurring, that he was fine to drive home. These dangerous attitudes still exist, particularly in less developed countries where alternatives like Uber – a ride hailing app – are not available.

Tangentially to safety, there are also salient moral dynamics to consider when we examine the working conditions in long-distance trucking. Truck drivers commonly face immense pressure to deliver their payloads within a set timeframe and face financial penalties if they do not. This often results in them foregoing regulated rest stops, forcing them to continue driving under conditions of extreme fatigue. Substituting truck drivers with autonomous trucks purports to solve these problems by obscuring them. The ethics of autonomous vehicles currently construed avoids asking these prior moral questions.

How then do we make an ethics of autonomous vehicles ethical, that is, responsible to and for the other? We could also phrase the question in terms that summarise this section's claims: What is our relationship to autonomous vehicles really like? Where can we locate the face, if at all, in autonomous vehicles? What does the compensatory logic of the ethics of autonomous vehicles, explored using the trolley problem framed within a utilitarian ethics, reveal about how ethical responsibility operates in such a framework? Before attempting to answer that, I turn now to how and what the (Levinasian) face signifies.

Meaning all by itself: The naked face

Levinas remarks that the 'face is meaning all by itself'. He also, in the same interview, says that the "face is signification, and signification without context" (1985; 86). By this Levinas means that ordinarily, who we are is signified within certain contexts – being a professor, a judge, the son of a famous politician, wearing certain clothes etc… But our alterity is not any one of these signifiers, or rather, our uniqueness amounts to something much more than the totality of these significations – our singularity, our face, is 'meaning all by itself'. Perpich (2008; 69) argues that the face

> represents the inadequacy of every image to the task of representing the other and, as such, paradoxically, represents the impossibility of its own representational activity.

Elsewhere, Perpich (194) captures the ambiguity and uncertainty that is at the heart of the face's signification: "'I am this', the face says, 'but not only this'. And even as it refuses representation in one sense, the face demands it in another".

Levinas (1985; 86) continues by arguing that the face "is the most naked [and…] the face is exposed, menaced". Exposed to, and menaced by what, or who? The self, who attempts to eliminate the ambiguity of the other's signification. By establishing the certainty of who or what the other is, the self seeks to affirms its own autonomy. The self then systematically erodes the alterity of the other in order to represent and categorise the other into profession, race, religion etc. However, the other resists my attempts to represent their alterity, by *facing* me in the face-to-face encounter. The Other resists their totalisation with a "resistance of what has no resistance – the

ethical resistance" (Levinas 1969; 199). It is this resistance that holds my subjectivity hostage and binds me with an infinite responsibility to and for the other.

What then does the face signify? Levinas (66) says that the face speaks. The manifestation of the face is already discourse. […] To present oneself by signifying [, by facing,] is to speak". Elsewhere, Levinas (1985; 89) expands on this: "the first word of the face is the 'Thou shalt not kill'". This imperative should not be understood only literally, but also as the metaphysical violence of reducing the other to the same. Levinas continues,

> there is a commandment in the appearance of the face, as if a master spoke to me. However, at the same time, the face of the Other is destitute, it is the poor for whom I can do all *(ibid*.).

The other thus arrests my freedom (to represent and categorise her) through the very force of their vulnerability.

How then can we not subject the other to the metaphysical violence of reducing them to the same? How can we demonstrate, in our everyday lives, our recognition of the irreducible singularity of the other and our infinite responsibility to safeguard that singularity? Levinas is frustratingly vague on this, and this is precisely because, as described in the first section, to do so would amount to 'constructing ethics', that is, prescribing a system of norms. Levinas (1998; 117) does however suggest that this is possible, and is epitomised in the slightest of everyday gestures, "the little there is, even the simple 'after you sir'". Perpich (2008; 135) clarifies Levinas' argument:

> [t]he events we recognize as ethical, from the polite gestures of social commerce to the selfless lives of saints, and everything in between, are

> predicated upon or find their condition in the unconditionality of being hostage [to the demands of the other].

The simple hand gesture that signifies to another to enter a building before oneself, and which demonstrates our recognition of the other's alterity, recalls Levinas's characterisation of the face as the mode, or *the way in which the other presents themselves*. What this means is that while the face might be 'meaning all by itself', it is not the holder of all the meaning the other is. The back, or the nape of the neck of the other, can signify the alterity of the other just as well as the face. Levinas (2001; 208), referencing a scene in Vasily Grossman's (1985) novel *Life and Fate*, talks about people lining up, in Moscow, to hear about news of friends and relatives arrested for political crimes and would "each [read] on the nape of the person in front of him the feelings and hopes of his misery". Here Levinas is interpreting Grossman (1985; 683 in Morgan 2007; 5) who describes one Yevgenia, who

> had never realized that the human back could be so expressive, could so vividly reflect a person's state of mind. People had a particular way of craning their necks as they came up to the windows; their backs, with their raised, tensed shoulders, seemed to be crying, to be sobbing and screaming.

The face is signification, but so too is the back and the neck; the face speaks of the suffering of the other and the concomitant demand to ease that suffering, but so too does the back and the neck.

In the previous section, I traced Toivakainen' s argument that the search for a machine-face is an example of compensatory logic, and that if the machine is to have a face, then it must bear its own ethical meaning. Toivakainen is sceptical about such a possibility. In light of this sections' circumscription of how and what the

face signifies, a better formulation is to ask 'how can the machine signify *its* alterity?' The answer seems straightforward enough: It cannot, because the machine has none to signify – it is not singular in the way a person is. What the machine can signify however, is the alterity of the other (singular) person and the responsibility that is owed to her. In other words, while the machine cannot signify *as* a face, that is, it cannot signify without context, it can signify the face of the other which does. The machine *announces or anticipates* the signification of the face which the self is infinitely responsible to[10].

Understanding the signification of machines in this way allows us to avoid the trap of a compensatory logic because the moral dynamics of Levinasian intersubjectivity – the self as for the other – remain front and centre regardless of the context, whether that be the elderly and care-bots, or the enhancement of road-safety and autonomous vehicles. In the final section of this paper, I turn to a more pedestrian (pun intended) technology than autonomous vehicles to demonstrate how a traffic signal might announce the signification of the face. In the process I argue that the technology of 'shared space' is a technology that can enable safer and more ethically responsible autonomous vehicles.

Before doing so, a summary of the section: Levinas says that the face signifies without context, but because the other is singular such a signification is always ambiguous and uncertain. Levinas also claims that it not only the face, but also the back and the neck which can signify the alterity of the other, or in another formulation, can 'speak' the other's alterity. I then argued that because a machine

[10] As a modality, the face might be understood as operating like a machine which produces alterity. The face then, could be a machine, but the machine could not be a face.

cannot be singular in the way a human can, it cannot signify as a face. The machine can however announce or anticipate the signification of the face.

Naked streets

Although there are no hard and fast definitions, the concept of 'shared space' "is an approach to street design which minimises demarcations between vehicles and pedestrians" (Moody & Melia 2013; 384). The concept originated with the Dutch traffic engineer Hans Monderman whose original aims were "to reduce accidents and congestion and to increase the flow of traffic" (*ibid.*). Hamilton-Baillie (2008; 133) argues that this can be achieved by the

> removal of the familiar characteristics associated with the highway, such as road markings, traffic signals, signs, kerbs, bollards and barriers [which] can dramatically change the relationship between people, places and traffic.

The removal of traffic lights and signals, inter alia, from certain streets has led to these also being called 'naked streets' (Mckone 2010). While the concept of shared space has evolved to include wider aims such as "a modal shift to walking and cycling, enhancement to the public realm and improved health" (Hamilton-Baillie 2008; 137), my focus will remain on the traffic safety aspect of shared space.

Hamilton-Baillie (133) outlines the rationale behind the idea of naked streets: "in the absence of rules, predictability and certainty, drivers have to rely on cultural signals and informal social protocols." In other words, ambiguity and uncertainty forces us to be more responsive through demanding greater attention from us within a particular traffic situation and hence, to take up greater responsibility. It is at this point where I

return to the notion of the signification of the face, whose core modality also operates through ambiguity and uncertainty. Monderman (in McNichol 2004) argues that

> a wide road with a lot of signs is telling a story. It's saying go ahead, don't worry, go as fast as you want, there's no need to pay attention to your surroundings.

That is to say, the face of the other is here anticipated by a wide road with lots of signs. The road and its signs, with all their pragmatic prescriptions, announces the responsibility the self owes to the other which impinges upon the agency of the self, in this case, to drive as fast as they please. The road here tells the story of a self who believes it can escape the unambiguous command of the face which says, 'don't kill me', in this case by speeding.

Another way the road might bear the meaning of the face of the other can be gleaned from Hamilton-Baillie (2004; 58) discussing Monderman in the context of footpaths:

> In a suburban street, a footpath meets a main estate road at right angles. In the United Kingdom, this insignificant intersection would probably be marked by a "safety barrier" to cut the footpath from the road. By contrast, Monderman expresses the presence of the footpath into the road through special surface treatment and a slight rise in level; not enough to constitute a speed hump, but enough to subtly draw the driver's attention to the status and significance of the footpath and the possible arrival of teenage skateboarders.

The slight rise in road level announces the face of the other who is in this instance a teenage skateboarder. The slight rise in the road level anticipates an unknowable

other whose uncertain appearance strips me of my agency to act unconditionally, which in this context amounts to driving in a carefree manner.

Monderman (in Jenkins 2016) argues that “the freedom to assess risk for ourselves is what makes us safer” and that “the greater the number of prescriptions, the more people’s sense of personal responsibility dwindles”. While the first half of Monderman’ s remark would seem to indicate a freedom only in service of the self, the second half reveals that such freedom is circumscribed by responsibility. This is Levinasian freedom – Levinas (in Goud 2008; 21) argues that the “willingness to serve [the Other] is actually freedom, namely election. In the place of autonomy, I put election based on untransferable responsibility”. Monderman (in Schulz 2006) also argues that “the many rules strip us of the most important thing: the ability to be considerate. We’re losing our capacity for socially responsible behaviour.” The ability to be considerate is exactly, like the simple gesture ‘after you, sir’, a demonstration of the investiture of our freedom as responsibility, that is, our infinite responsibility to and for the other.

One conference, dedicated to implementing Monderman’ s ideas of naked streets, adopted the guiding motto ‘Unsafe is safe’ (Schulz 2006). This recalls the modality of the Levinasian face – as that which resists (its totalisation) without resistance. It is now an easy step to arguing for this paper’s claim: roads can be made safer by making them signify the ambiguity and uncertainty of the faces that traverse them. The road in its nakedness announces the nakedness of the face of other who is in the instance a road-user. Naked streets and shared space demonstrate how the notion of an infinite responsibility toward an unknowable other who can only signify their alterity ambiguously and uncertainly might be feasible in practice.

Conclusion

In response to Toivakainen' s invocation of Turkle's (2011; 283 in Toivakainen 2016; 280) remark that certain technologies might "obscure a problem by 'solving' it without addressing it", I asked 'how might autonomous vehicles obscure the problem of traffic safety by solving it without addressing it?' One way the ethics of autonomous vehicles tries to solve the problem of autonomous vehicles is to ask whose safety should be prioritised in the event of an accident involving an autonomous vehicle. Determining this involves categorising road-users into, inter alia, 'pedestrian' or 'passenger'. The rules of the road and other traffic signals assist in this demarcation and are meant to remove any ambiguity around such demarcation. In trying to remove the uncertainty such distinctions entail we try to solve the problem of who should be assigned greater weight, the pedestrian or the passenger, in the moral calculations required by the trolley problem. In doing so we do not address the problem autonomous vehicles present to our notion of responsibility. Not accountability, or blameworthiness, but our *infinite responsibility* to and for the other which must be worked out in the finite confines of the vehicles, autonomous and otherwise, that transport us.

If, as this paper argues, we understand autonomous vehicles as announcements or anticipations of the face of the other, then the ethics of autonomous vehicles can escape the compensatory logic it has fallen into through its fixation on the trolley problem as a way to understand the ethical stakes. To do so, autonomous vehicles will need to embrace the ambiguity and uncertainty that the face signifies, as the technology of naked streets tries to do. In risking less safe autonomous vehicles, we might make roads, and road-users, safer in the process.

References

Atterton, P. (2011). ‘Levinas and Our Moral Responsibility Toward Other Animals’, *Inquiry*, 54(6), 633-649.

Awad, E., Dsouza,S. , Kim, R., Schulz, J. ,Henrich,J. ,Shariff,A. , Bonnefon, J & Rahwan, I. (2018). ‘The Moral Machine experiment’, *Nature*, 563, 59–64.

Bauman, Z. (1993). *Postmodern Ethics*. Oxford: Blackwell

Casey, E.S. (2001). ‘Taking a Glance at the Environment: Prolegomena to an Ethics of the Environment’, *Research in Phenomenology,*31(1), 1-21.

Derrida, J. (1978). ‘Violence and Metaphysics’ in *Writing and Difference.* Trans. By A. Bass. London: Routledge, pp 97-192.

Derrida, J. (2008). *The Animal That Therefore I Am,* in Mallet, M.L.(Ed.). Trans by D. Wills. Bronx, NY: Fordham University Press.

Diehm, C. (2000). ‘Facing nature: Levinas beyond the human’, *Philosophy Today*, 44(1), 51-59.

Floridi, L. (2002). ‘On the intrinsic value of information objects and the info-sphere’. *Ethics and Information Technology*, 4, 287–304.

Hamilton-Baillie, B. (2004).’Urban design: Why don't we do it in the road? Modifying traffic behavior through legible urban design, *Journal of Urban Technology*, 11(1), 43-62.

Hamilton-Baillie, B. (2008). ‘Towards shared space’, *Urban Design International*, 13(2), 130-138.

Hevelke, A., & Nida-Rümelin, J. (2014). ‘Responsibility for crashes of autonomous vehicles: An ethical analysis’. *Science and Engineering Ethics*, 21(3), 619–630.

Himmelreich, J. (2018). ‘Never Mind the Trolley: The Ethics of Autonomous Vehicles in Mundane Situations’, *Ethical Theory and Moral Practice*, 21, 669–684.

Hogan, K. (2017). ‘Is the machine question the same question as the animal question? *Ethics and Information Technology*, 19, 29-38.

Foot, P. (1967). ‘The problem of abortion and the doctrine of double effect’, *Oxford Review*, 5, 5–15.

Gogoll, J., & Muller, J. F. (2017). ‘Autonomous cars: In favor of a mandatory ethics setting’, *Science and Engineering Ethics*, 23(3), 681–700.

Goodall, N.J. (2016). ‘Away from trolley problems and toward risk management’, *Applied Artificial Intelligence*, 30(8),810–821.

Goud, J.F. (2008). ‘What One Asks of Oneself, One Asks of a Saint: A Dialogue with Emmanuel Levinas, 1980-1981.’ *Levinas Studies*, 3, Pittsburgh, Pennsylvania: Duquesne University Press,1-34.

Grossman, V. (1985). *Life and Fate*. Trans. Chandler, R. New York: Harper and Row.

Gunkel, J. D. (2012). *The Machine Question*. Cambridge, MA: The MIT Press.

Jenkins, S. (2016). ‘The removal of road markings is to be celebrated. We are safer without them’, *The Guardian*, 4 February, available at https://www.theguardian.com/commentisfree/2016/feb/04/removal-road-markings-safer-fewer-accidents-drivers (accessed 6 August 2019).

Introna, L. (2014). ‘Ethics and the flesh: Being touched by the otherness of things’. In O. Petursdottir (Ed.), *Ruin memories: Material, aesthetics and the archeology of the recent past*. Oxford: Routledge. 41-61

Levinas, E. (1969). *Totality and Infinity: An Essay on Exteriority*. Trans. Lingis, A. Pittsburgh, PA: Duquesne University Press

– (1985). *Ethics and Infinity: Conversations with Philippe Nemo*. Trans. by R.A. Cohen. Pittsburgh: Duquesne University Press.

– (1988). 'The Paradox of Morality': An Interview with Emmanuel Levinas by T. Wright, P. Hayes and A. Ainley, Trans. A. Benjamin and T. Wright, in *The Provocation of Levinas: Rethinking the Other*, ed. R. Bernasconi & D. Wood. London: Routledge. 160-180

– (1989). *The Levinas Reader*. Ed S. Hand. Oxford: Blackwell Publishers.

– (1998). O*therwise Than Being, or Beyond Essence.* Trans. A. Lingis. Dordrecht: Kluwer Academic Publishers.

– (2001). *Is it Righteous to Be? Interviews with Emmanuel Levinas*. Ed. Robbins, J. Stanford, CA: Stanford University Press

Llewelyn, J. (1991). 'Am I Obsessed by Bobby? (Humanism of the Other Animal)' in *Re-Reading Levinas*, Bernasconi, R. & Critchley, S. (eds.), 234-245. Bloomington: Indiana University Press.

McNichol, (2004). 'Roads Gone Wild', *Wired*, 1 December, available at https://www.wired.com/2004/12/traffic/ (accessed 6 August 2019).

McKone, J. (2010). '"Naked Streets" without traffic light improve flow and safety', *The City Fix*, 18 October, available at https://thecityfix.com/blog/naked-streets-without-traffic-lights-improve-flow-and-safety/ (accessed 6 August 2019).

Moody, S., and Melia, S. (2013). 'Shared space – Research, policy and problems', *Proceedings of the Institution of Civil Engineers*, 167(6), 384–392.

Morgan, M.L. (2007). *Discovering Levinas*. Cambridge: Cambridge University Press.

Nyholm S. & Smids J. (2016). 'The ethics of accident-algorithms for self-driving cars: an applied trolley problem?', *Ethical Theory and Moral Practice*, 19, 1275–1289.

Perpich, D. (2008). *The Ethics of Emmanuel Levinas*. Stanford, California: Stanford University Press.

Schutz, M. (2006). 'Controlled Chaos: European Cities do away with Traffic Signs', *Spiegel Online*, 16 November, available at https://www.spiegel.de/international/spiegel/controlled-chaos-european-cities-do-away-with-traffic-signs-a-448747.html (accessed 6 August 2019)

Toivakainen, N. (2016). 'Machines and the face of ethics', *Ethics and Information Technology*, 18, 269-282.

Turkle, S. (2011). *Alone together*. New York, NY: Basic Books.